Journal of Parkinson's Disease 11 (2021) 395–404
DOI 10.3233/JPD-202472
IOS Press

Systematic Review

Parkinson's Disease Subtypes: Critical Appraisal and Recommendations

Tiago A. Mestre[a,b,*], Seyed-Mohammad Fereshtehnejad[b], Daniela Berg[c], Nicolaas I. Bohnen[d], Kathy Dujardin[e], Roberto Erro[f], Alberto J. Espay[g], Glenda Halliday[h], Jacobus J. van Hilten[i], Michele T. Hu[j], Beomseok Jeon[k], Christine Klein[l], Albert F.G. Leentjens[m], Johan Marinus[n], Brit Mollenhauer[o], Ronald Postuma[p], Rajasumi Rajalingam[q], Mayela Rodríguez-Violante[r], Tanya Simuni[s], D. James Surmeier[t], Daniel Weintraub[u], Michael P. McDermott[v], Michael Lawton[w] and Connie Marras[q]

[a]*Parkinson's disease and Movement Disorders Center, Division of Neurology, Department of Medicine, The Ottawa Hospital Research Institute, The University of Ottawa Brain and Research Institute, Ottawa, ON, Canada*
[b]*Division of Neurology, Department of Medicine, The Ottawa Hospital, University of Ottawa, Ottawa, ON, Canada*
[c]*Department of Neurology, Christian-Albrechts-University, Kiel, Germany*
[d]*Departments of Radiology & Neurology, University of Michigan, University of Michigan Udall Center, Ann Arbor VAMC, Ann Arbor, MI, USA*
[e]*Movement Disorders Department, Center of Excellence for Neurodegenerative Diseases LiCEND, Lille, France*
[f]*Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", Neuroscience Section, University of Salerno, Baronissi (SA), Italy*
[g]*James J. and Joan A. Gardner Family Center for Parkinson's Disease and Movement Disorders, Department of Neurology, University of Cincinnati, Cincinnati, OH, USA*
[h]*Brain and Mind Centre and Central Clinical School, Faculty of Medicine and Health, University of Sydney, Australia*
[i]*Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands*
[j]*Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Neurology Department, Oxford, United Kingdom*
[k]*Department of Neurology, Seoul National University Hospital, Seoul, Korea*
[l]*Institute of Neurogenetics, University of Luebeck, Luebeck, Germany*
[m]*Department of Psychiatry, Maastricht University Medical Center, Maastricht, The Netherlands*
[n]*Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands*
[o]*Paracelsus-Elena-Klinik, Kassel and University Medical Center Goettingen, Department of Neurology, Kassel, Germany*
[p]*Department of Neurology, McGill University, Montreal, Quebec, Canada*
[q]*Edmond J. Safra Program in Parkinson's Disease and the Morton and Gloria Shulman Movement Disorders Clinic, Toronto Western Hospital, University Health Network, Toronto, Canada*
[r]*National Institute of Neurology and Neurosurgery, Mexico City, Mexico*
[s]*Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*
[t]*Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*

*Correspondence to: Tiago A. Mestre, MD, MSc, The Ottawa Hospital, Civic Campus, 1053 Carling Ave, Rm C2196, Ottawa ON K1Y 4E9, Canada. Tel.: +1 613 798 5555 /Ext. 18986; Fax: +1 613 761 4507; E-mail: tmestre@toh.ca.

ISSN 1877-7171

[u]*Departments of Psychiatry and Neurology, Perelman School of Medicine at the University of Pennsylvania; Parkinson's Disease Research, Education and Clinical Center (PADRECC), Philadelphia Veterans Affairs Medical Center, Philadelphia, PA, USA*
[v]*Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA*
[w]*Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom*

Accepted 4 February 2021
Pre-press 5 March 2021

Abstract.
Background: In Parkinson's disease (PD), there is heterogeneity in the clinical presentation and underlying biology. Research on PD subtypes aims to understand this heterogeneity with potential contribution for the knowledge of disease pathophysiology, natural history and therapeutic development. There have been many studies of PD subtypes but their impact remains unclear with limited application in research or clinical practice.
Objective: To critically evaluate PD subtyping systems.
Methods: We conducted a systematic review of PD subtypes, assessing the characteristics of the studies reporting a subtyping system for the first time. We completed a critical appraisal of their methodologic quality and clinical applicability using standardized checklists.
Results: We included 38 studies. The majority were cross-sectional ($n = 26$, 68.4%), used a data-driven approach ($n = 25$, 65.8%), and non-clinical biomarkers were rarely used ($n = 5$, 13.1%). Motor characteristics were the domain most commonly reported to differentiate PD subtypes. Most of the studies did not achieve the top rating across items of a Methodologic Quality checklist. In a Clinical Applicability Checklist, the clinical importance of differences between subtypes, potential treatment implications and applicability to the general population were rated poorly, and subtype stability over time and prognostic value were largely unknown.
Conclusion: Subtyping studies undertaken to date have significant methodologic shortcomings and most have questionable clinical applicability and unknown biological relevance. The clinical and biological signature of PD may be unique to the individual, rendering PD resistant to meaningful cluster solutions. New approaches that acknowledge the individual-level heterogeneity and that are more aligned with personalized medicine are needed.

Keywords: Parkinson's disease, heterogeneity, subtypes

INTRODUCTION

Parkinson's disease (PD) is known for its heterogeneity in terms of both clinical practice and underlying biology [1]. One approach to understanding the variability present in PD is to identify subtypes; that is, to define groups of patients with a set of differentiating features at a clinical, biological, genetic or pathological level.

Research on PD subtypes has garnered considerable attention for its potential to elucidate the disease pathophysiology and natural history and ultimately, guide therapeutic development. However, the impact of PD subtype identification on our understanding of PD pathogenesis or clinical treatment remains unclear with limited application in research or clinical practice. Methodologic heterogeneity, lack of reproducibility, poor prognostic value, and limited clinical applicability are among the possible reasons for the reduced impact of PD subtypes [2].

The Task Force for PD Subtypes was convened in 2018 by the International Parkinson's disease and Movement Disorders Society (MDS) to critically appraise the available PD subtyping studies and to provide guidance for the design and conduct of future studies on PD subtypes. In this systematic review of PD subtypes, we evaluate both the methodologic quality and clinical applicability of studies proposing a new subtyping system. Based on these findings, we provide our assessment of the future of subtyping research.

METHODS

Overview

We searched for PD subtyping studies in PubMed/MEDLINE using the following search terms: 'Parkinson Disease'[Mesh] AND ('Subtyp*' OR 'Phenotype'[Mesh] OR 'Phenotyp*' OR 'Biomarkers'

[Mesh] OR 'Clinical Feature*' OR 'endophenotyp*'). We defined a PD subtyping study as any research study conducted with the purpose of dividing PD patients into subtypes, as stated by its authors, or identified distinct groups of PD patients that were discussed as possible subtypes. We excluded studies that focused on a subgroup of PD patients, such as those with causative or risk-associated genetic variants. We only included the initial report of a given PD subtype classification system. The methodologic quality of each study and the clinical applicability of each identified PD subtype system were evaluated using a standardized approach. Pairs of reviewers abstracted data from the included studies, including study design, baseline characteristics, PD subtyping methodology and results. We assessed the methodologic quality and clinical applicability of the included studies using two tools developed for the current study: a 13-item checklist for Methodologic Quality (item score range: 0–2, higher score being better, see Table 4) and an 11-item checklist for Clinical Applicability (items rated as Unknown, Limited/Low, Moderate, and Strong, see Supplementary Table 1). We compared rating frequencies between two publication periods (1980–2014 vs. 2015–2019) and two subtyping methodologic approaches (data-driven vs. hypothesis-driven) using Fisher's exact tests. The definition of publication periods was pragmatic to allow for a balanced distribution of included studies into a more recent group (representing the current state of the field) and older studies allowing for a sufficient sample size in each group to test for temporal trends. A detailed description of Methods is provided in the Supplementary Material.

RESULTS

We identified 51 studies after the initial screening, of which 13 studies were excluded after full-text review (Supplementary Figure 1). Of the 38 included studies, 68.4% were cross-sectional ($n = 26$) and 84.2% were conducted in a tertiary care center ($n = 32$). Across all studies, the mean disease duration was 59.8 months. Nine studies (23.7%) included untreated participants exclusively. Commonly used descriptors of PD patients (e.g., Unified Parkinson's Disease Rating Scale scores) were not reported in a significant number of studies (Table 1).

Table 1
Study design and sample characteristics of included studies

	$n = 38$
STUDY CHARACTERISTICS	
Sample size, mean (range)	293 (15–1601)
Setting - N (%)	
Single-center	21 (55.3)
Multi-center	15 (39.5)
Not reported	2 (5.2)
Recruitment Source - N (%)	
Tertiary care	32 (84.2)
Community-based/Tertiary Care	2 (5.2)
Not reported	4 (10.5)
Design - N (%)	
Cross-sectional	26 (68.4)
Longitudinal	12 (31.6)
Subtyping Approach - N (%)	
Hypothesis-driven	8 (21.0)
Data-driven	25 (65.8)
Hypothesis- and Data-driven	2 (5.3)
Not reported	3 (7.9)
PD Diagnosis - N (%)	
By Neurologist	28 (73.7)
Not specified	10 (26.3)
PD Diagnostic criteria - N (%)	
UK Brain Bank Criteria	22 (57.9)
UK Brain Bank Criteria+DAT Scan	1 (2.6)
Other formal criteria	2 (5.2)
No formal criteria/Investigator opinion	7 (18.4)
Not reported	6 (15.8)
SAMPLE CHARACTERISTICS*	
Age, mean (range)	64.9 (57.5–70.6)
	Not reported = 5
Male (proportion), mean (range)	62.6 (37.1–66.8)
	Not reported = 8
Disease Duration (months), mean (range)	47.4 (6.5–121.9)
	Not reported = 8
Dopaminergic treatment status (at baseline) - N (%)	
Untreated	9 (23.7)
Treated	9 (23.7)
Mixed	13 (34.2)
Not reported	7 (18.4)

*Means are weighted by study sample size.

Identification and description of reported PD subtypes

In the 38 included studies, 65.8% ($n = 25$) used exclusively a data-driven approach and 21.1% ($n = 8$) used exclusively a hypothesis-driven approach, evaluating differences between pre-determined groups. Five studies either used a combined approach ($n = 2$) or alternative analytical approaches such as regression analyses or *post-hoc* grouping based on results of the study ($n = 1$). Most of the data-driven studies ($n = 16/25$, 64.0%) used at least three phenotypic domains to identify PD subtypes, while 7/8 hypothesis-driven studies used a single domain. Overall, the motor domain was most frequently used

Table 2
Phenotypic domains and statistical methods used in the included studies

	Hypothesis-driven studies $n=8$	Data-driven studies $n=25$
Number of phenotypic domains used for subtyping - N (%)		
Single-domain	7 (87.5)	9 (36.0)
Two domains	1 (12.5)	-
>= 3 domains		16 (64.0)
Phenotypic domain(s) - N (%)		
Demographic	2 (25.0)	8 (32.0)
Motor	3 (37.5)	18 (72.0)
Cognitive	1 (12.5)	17 (68.0)
Emotional	1 (12.5)	17 (68.0)
Autonomic	1 (12.5)	11 (44.0)
Treatment	-	3 (12.0)
Non-clinical Biomarkers*	-	5 (20.0)
Imaging (n = 4), biochemical (n = 1)		
Statistical approaches		
Pre-determined groups	8 (100)	-
Hierarchical cluster analysis	-	7 (28.0)
Non-hierarchical cluster analysis	-	16 (64.0)
Hierarchical and Non-hierarchical cluster analysis		1 (4.0)
Other		1 (4.0)

Five studies were not included in this analysis for the following reasons: combined hypothesis- and data-driven approach ($n=2$) and other methods (regression analyses, $n=2$; subtype criteria defined *post-hoc*, $n=1$).

for subtyping in both types of studies, followed by non-motor domains, such as cognitive, emotional or autonomic, in studies using a data-driven approach, and the demographic domain (age or sex) in hypothesis-driven studies. Only five studies included non-clinical biomarkers, all of which adopted a data-driven approach (Table 2).

From the descriptions of the resulting subtypes, reviewers identified which specific features distinguished the groups in a statistically significant manner. The reviewed PD subtypes were found to have on average three statistically significant distinctive features between subtypes (mean = 3.3, range: 0–9). Features within the motor domain were the most commonly reported to differentiate PD subtypes. Imaging biomarkers were found to be a differentiating feature in three out of four studies that included this domain (Table 3).

Methodologic quality (Table 4, Supplementary Table 2)

Most of the studies did not achieve the top rating across methodologic quality items, with the exception

Table 3
Phenotypic domains discriminating PD subtypes reported in the included studies. We included both variables initially used to identify PD subtypes and additional variables found to be statistically significantly different between subtypes after the identification of PD subtypes in a study

Phenotypic domain	n	%
Demographic	13	11.8%
Motor	35	31.8%
Cognitive	17	15.4%
Emotional	13	11.8%
Autonomic	5	4.6%
Treatment	4	3.6%
Other non-motor	5	4.6%
Quality of Life/Functional ability	6	5.4%
Time-defined measures*	9	8.2%
Non-clinical biomarkers**	3	2.7%

*Examples: disease duration, UPDRS/disease duration. **Neuroimaging only.

of the items 'diagnostic methods' and 'variables compared between subtypes'. Statistical methods were difficult to evaluate. Emerging themes were recruitment from clinics rather than community sources, the lack of reporting of specific methods of clustering, lack of specification of how the number of clusters was determined, lack of justification of sample size, lack of adjustment for multiple comparisons when comparing features of the clusters, and lack of adjustment for or exploration in the report of the impact of fundamental baseline characteristics such as disease duration [3].

The four studies with the best quality ratings [4–7] were multi-center, longitudinal, data-driven studies published after 2016. These studies used more than one clinical domain (motor and non-motor) and only one study used additional CSF and neuroimaging biomarker data [7]. Two of the four studies used a homogeneous PD population in terms of stage/disease duration and conducted validation of the identified PD subtypes [4, 5]. In spite of incorporating similar clinical domains represented in the cluster analyses, there were no clear similarities among subtypes described in these two studies [4, 5]. Only one study developed an algorithm to classify individual participants into a PD subtype [4] and only one of these studies assessed the temporal stability of PD subtypes [6].

Compared with hypothesis-driven studies, data-driven studies more frequently used multi-center data collection ($p=0.04$) and more than one clinical domain or biomarker ($p=0.003$). We did not identify significant differences between the groups of studies published before or after 2015 (Supplementary Table 4).

Table 4
Item score distributions in the 13-item Methodologic Quality tool of included subtyping studies using a data-driven or hypothesis-driven approach

Item	Score rating	Hypothesis-driven ($n=8$)	Data-driven ($n=25$)	p
Disease stages/duration (study population)	0 = mixture of stages/disease duration at baseline or not reported	5 (62.5)	18 (72)	0.67
	1 = homogeneous disease stage/duration	3 (37.5)	7 (28)	
Study setting (*representativeness*)	0 = single-center or not reported	7 (87.5)	10 (40)	0.04
	1 = multi-center	1 (12.5)	15 (60)	
Recruitment source (generalizability)	0 = clinic-based or not reported	8 (100)	25 (100)	1
	1 = community or population-based	0 (0)	0 (0)	
Diagnostic methods	0 = not described or 1 or 2 not applicable	2 (25)	1 (4)	0.04
	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	6 (75)	24 (96)	
	2 = *postmortem* diagnosis	0 (0)	0 (0)	
Sampling method	0 = convenience or not reported	5 (62.5)	21 (84)	0.32
	1 = consecutive or random	3 (37.5)	4 (16)	
Comprehensiveness of data used for subtyping (subtype definition)	0 = single clinical or biomarker domain	8 (100)	9 (36)	0.003
	1 = > 1 clinical domains or biomarkers	0 (0)	16 (64)	
Variables compared between subtypes (*post hoc*)	0 = not done	0 (0)	2 (8)	1
	1 = single clinical domain or biomarker	0 (0)	0 (0)	
	2 = > 1 clinical domains or biomarkers	8 (100)	23 (92)	
Statistical methods used for subtyping	0 = low quality	4 (50)	4 (16)	0.11
	1 = intermediate quality	3 (37.5)	10 (40)	
	2 = high quality	1 (12.5)	11 (44)	
Longitudinal follow-up	0 = none (cross-sectional) or longitudinal < 1 year	6 (75)	17 (68)	0.68
	1 = short-term (1–3 years) OR longer-term but < 3 time-points	2 (25)	4 (16)	
	2 = longer-term (> 3 year) AND > = 3 time-points	0 (0)	4 (16)	
Completeness of follow-up	0 = cross-sectional or ≤ 50% complete or not reported	7 (87.5)	20 (80)	1
	1 = 50–75% complete	1 (12.5)	2 (8)	
	2 = > 75 % complete	0 (0)	3 (12)	
Subtype stability	0 = not assessed	8 (100)	23 (92)	1
	1 = assessed	0 (0)	2 (8)	
Algorithm for classifying individual patients	0 = not provided	0 (0)	24 (96)	< 0.001
	1 = provided	8 (100)	1 (4)	
Validation (internal or external)	0 = not assessed	8 (100)	19 (76)	0.57
	1 = use of a test set from the same population	0 (0)	4 (16)	
	2 = validation in an external population	0 (0)	2 (8)	

Clinical applicability (Supplementary Table 3, Supplementary Figure 2)

Overall, items reflecting the clinical importance of differences between subtypes, potential treatment implications and applicability to the general population of PD patients were rated poorly for most studies. Subtype stability over time and prognostic value were largely unknown due to the paucity of longitudinal studies. Compared with hypothesis-driven subtypes, data-driven subtyping was seen as burdensome and time-consuming ($p=0.01$). Both data-driven and hypothesis-driven subtyping were usually rated as inexpensive. There were no clear differences between newer and older studies for the different items of the Clinical Applicability tool (Supplementary Figure 3)

DISCUSSION

Our systematic review has revealed gaps in the field and highlighted the limitations of current approaches to PD subtyping. It thereby informs recommendations on the methodology of future subtyping studies and suggests alternative directions. Recommendations are highlighted in italics.

Research methodology: Can we do better?

Quality ratings for subtyping studies revealed clear areas for improvement. More extensive use of longitudinal data is critical for an understanding of the stability of proposed subtypes, and their prognostic value. The use of longitudinal data to define or evaluate subtypes does appear to be more common in the last 5 years [4–8], making use of large, publicly available cohorts [4, 9]. To our knowledge, only one study has used longitudinal profiling as the basis for defining subtypes [7], incorporating data on the evolution of clinical or biological features across time into the definition of subtypes. Alternatively, serial cluster analyses could provide data about the stability of proposed subtypes and the influence of disease duration on their characteristics. Such approaches could provide additional prognostic value, using information about the early evolution of disease to inform later prognosis or underlying biology.

An area for future study is incorporating the prodromal phase of PD, which presents the opportunity to start defining subtypes of the disease earlier in the pathological process, which may provide new insights. A better understanding of the starting point and initial progression may better predict the course and subsequent clinical progression and offer the opportunity for an early target-specific therapeutic development.

A recurring reporting flaw was failure to describe statistical methods in detail. This issue may become more frequent as machine learning will probably play an important role in complex (hypothesis-free) analyses, particularly as the scope of available data expands [10]. It is important that the clustering method that is used is explicitly described in sufficient detail to facilitate independent replication and pursuit of subsequent hypothesis-driven studies [11].

Data-driven analyses were rarely replicated in a separate cohort. This is likely related to the fact that similar data are not always available on different cohorts and cohorts differ in their eligibility criteria (and assessments), resulting in fundamentally different characteristics of patients. However, it was also rare for studies to use internal replication techniques such as resampling. Internal or external replication attempts are important to assess the general applicability of suggested subtyping methods.

It was noted that the domains covered in subtyping studies and the instruments used were variable, limiting the value of comparisons across studies. Consensus moving forward as to the methodologic approach and the key core set of domains would be helpful for the field.

Subtyping by design

Possible uses of subtyping include clinical prognostication or identifying biological subtypes that predict therapeutic response to symptomatic (targeting convergent mechanisms, such as dopamine deficiency) or disease-modifying interventions (targeting divergent biological mechanisms, such as mitochondrial dysfunction). Different study designs may serve different purposes.

Design of subtyping studies fell into two main categories: 1) data-driven analyses and 2) hypothesis-driven analyses based on pre-determined groups. Hypothesis-driven studies were usually single domain analyses of a clinical or demographic characteristic such as age of onset, or tremor-dominant vs. akinetic-rigid ('motor subtyping'), and this was felt to be an inferior approach compared with considering multiple domains of data in a hypothesis-free subtyping exercise. However, it may be that focused hypothesis-driven evaluation of pre-defined groups is important for answering specific research questions, and this approach may have more potential for direct neurobiological mapping with concomitant biomarker measurement. For example, this approach has successfully identified biological differences between individuals based on motor phenotype [12, 13]. Findings from data-driven studies should be similarly used for hypothesis-driven studies to define the clinical applicability and/or biological underpinnings of the described subtypes. Importantly, clinical applicability may not be a relevant objective of some subtyping efforts.

Subtyping studies have usually enrolled individuals at various points in the disease course and analyzed the cohort without stratification on disease duration or stage. This approach makes it more difficult to describe phenotypic variability at specific phases of the disease. In addition, patients may be misclassified if individuals at different disease durations are included without taking the phenotypic changes with disease duration into account. Certain aspects such as motor phenotype tend to be more heterogeneous early in the disease process, converging to a common phenotype toward the end [14, 15]. Thus, subtyping exercises may produce very different results depending on the distribution of disease duration or stage at which the subtyping is performed. We would recommend subtyping based on

longitudinal data starting from a defined disease duration or milestone in order to maximize interpretability and minimize these confounding issues.

The included studies were all clinic-based, rather than recruiting from community sources. This may be a more or less important source of bias depending on the purpose of the subtyping; individuals from these two sources are unlikely to differ in the underlying biology of the disease but may well differ in clinical severity and prognosis. A community-based sample would be more generalizable.

What is the clinical and biological relevance of subtypes?

The clinical importance of the differences between the defined subtypes was deemed moderate or less for the vast majority of studies. Whether or not the proposed subtypes are stable, or become more or less different over time, could not be evaluated in this review for most classification systems since longitudinal differences were not assessed in the original descriptions of the subtypes. These limitations raise important questions about the clinical relevance of subtypes defined to date, with the caveat that we have not evaluated the literature following up the initial descriptions of the subtypes. We are aware of many studies evaluating subtypes defined by motor features contrasting tremor-dominant (TD) with other phenotypes. Some have looked at the prognostic implications of these subtypes [16, 17] and several studies have demonstrated that a high proportion of individuals with the TD phenotype will switch to a PIGD phenotype over time [14, 15, 18]. An adverse cognitive prognosis and prominent cortical Lewy body involvement associated with the non-tremor dominant phenotypes confirms the prognostic relevance of those subtypes [19]. Recently, there has also been an independent examination of the clinical and pathological evolution of data-driven subtypes as defined by an earlier cohort study [4]. Important differences in time to important clinical milestones between the subgroups were shown in a separate cohort, supporting the clinical importance of those subtypes [20]. Unfortunately, such studies are rare and we encourage such prognostic studies to clarify the clinical relevance of any described subtypes.

The relationship between clinically-based subtypes and the underlying biology of disease has rarely been assessed to date. Several studies have evaluated the biological characteristics of TD and PIGD phenotypes, finding differences in CSF composition (alpha-synuclein [21], 5-hydroxyindoleacetic acid, glycine [12]) and degree of cardiac sympathetic denervation, for example [21]. We are aware of only two studies evaluating the biomarker profile of data-driven subtypes, identifying a possibly pro-inflammatory profile [9] and an "Alzheimer's-like" profile of low CSF amyloid-β and high tau [4] associated with a severe motor and non-motor disease subtype. In the absence of validated non-clinical biomarkers of PD progression, the value of clinically-based subtyping to help direct disease-modifying therapies targeting specific biological processes is unknown. Establishing the relationship between clinically-based subtypes and biomarkers of the underlying disease biology is an important goal, as it may inform the development of subtype-targeted therapies.

We also found that there was little inclusion of biomarkers in studies defining PD subtypes. This has the same implications as mentioned above, limiting our understanding of the biological relevance of subtypes. We further recommend biologically-based subtyping studies.

Can subtyping be incorporated into clinical practice?

Clinical applicability may not be the goal of all subtyping efforts, such as those seeking to understand heterogeneity in the underlying biology of the disease. Therefore, clinical applicability may not be a relevant criterion on which to rate some studies. Nonetheless, it is relevant to note that our expert group had reservations about the feasibility of subtyping as part of routine clinical practice with the tools currently available. That conclusion was driven by 1) the extra time required to assign individuals to subtypes using multi-domain data, beyond that allocated to usual clinic visits and 2) the lack of an algorithm to classify individual patients. An advantage of hypothesis-based single-domain subtyping systems is the availability of a clear and simple algorithm to assign individuals to groups, facilitating both replication of findings in different cohorts [20] and application in a clinical setting. It is rare that data-driven studies use their results to derive an algorithm for assigning individuals to subtypes, although there is at least one exception [4, 20]. Even if feasible, the applicability of the results to individuals is unclear given that subtyping has so far been studied only at the group level. Studies examining the prognostic or biological relevance at the individual level are needed to

establish clinical applicability. Furthermore, simple classification systems based on a manageable number of variables would be a prerequisite for adoption by the clinical community; however, simplicity is at odds with the fact that PD is etiologically and phenotypically complex, and being a disease associated with aging, subject to multiple intersecting biological processes [22] and co-existing medical comorbidities [23] that can intervene to influence the phenotype at any point in the disease process.

Is subtyping possible?

Subtyping studies undertaken to date have significant methodologic shortcomings and most have questionable clinical applicability. Many studies were limited in terms of the clinical domains considered to define subtypes, and measures used in data-driven analyses were highly heterogeneous, resulting in highly variable findings. Even the two data-driven studies receiving the highest quality ratings and incorporating variables from comparable domains showed limited similarity of the subtypes they found. Despite decades of subtyping research, there has been minimal incorporation of subtype information into research or formal incorporation in clinical practice. These observations call into question the feasibility of clinical subtyping and suggest that alternative approaches to describing and understanding the heterogeneity of PD are needed.

As described above, PD is biologically complex, a result of many intersecting processes that vary from person to person. Blood and cerebrospinal fluid parameters associated with PD show high inter-individual heterogeneity, requiring combinations of multiple markers to optimize diagnostic accuracy [24]. Variable co-pathology adds to the heterogeneity [25]. Several 'causative' genes, none completely penetrant, and a vast array of susceptibility genes governing widely-varying metabolic processes have been elucidated [26]. RNA expression studies demonstrate the complex genetic and biologic heterogeneity of PD [27]. It is a testament to this complexity that metabolomics are now being used to measure the downstream effects of a vast array of contributory genetic, environmental and physiological processes [28]. As a result, the phenotype of PD is unique to an individual. Given these challenges, it seems unlikely that a purely clinically-based subtyping system measuring early disease features and seeking to place patients into a small number of categories will be able to adequately describe PD heterogeneity in order to provide accurate pathophysiological or prognostic insights. It remains to be seen whether this can be improved by the discovery of better biomarkers but underlying biological heterogeneity may also render biomarker-based subtyping resistant to meaningful cluster solutions. This idea is not new; it has been previously proposed that PD is not one disease but rather a syndrome representing the manifestations of multiple or even an infinite number of underlying diseases [29].

Contemporary medicine is increasingly promoting personalized treatment, including in Parkinson's disease, with a focus on the individual [30]. To date, subtyping places individuals in groups with similar but not identical features. This may represent an important step toward identifying individuals that can respond preferentially to certain treatments but by virtue of placing individuals within a group will inevitably fall short of the truly 'personal' goal. Although perhaps more challenging, modern computational techniques are increasingly allowing us to manage vast amounts of data and may soon allow us to take full advantage of the information available in the heterogeneity to describe individuals without the need for group-level subtyping. Many of the recommendations outlined above could apply to future studies where the unit of measure is the individual's disease fingerprint rather than the group phenotype. Granted, such an individual approach poses financial and logistical challenges when it comes to clinical trials which will have to be overcome. Nonetheless, having reviewed the existing literature on subtyping and explored the methodologic pitfalls and challenges associated with performing the optimal subtyping studies described above, it is time to re-evaluate our approach to understanding and describing PD heterogeneity.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the International Parkinson and Movement Disorders Society (MDS) for facilitating meetings of the Task Force and, particularly, the assistance of Sarah Smith from the MDS for assistance in coordination and communications.

CONFLICT OF INTEREST

Michele Hu reports that ODPC Discovery COhort was funded by the Monument Trust Discovery Award

from Parkinson's UK and supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford and the Dementias and Neurodegenerative Diseases Research Network (DeNDRoN).

Nicholas Bohnen received research funding from the NIH, Department of Veterans Affairs and the Michael J. Fox Foundation.

Glenda Halliday is a National Health and Medical Research Council of Australia Leadership Fellow (#1176607).

The remaining authors report there are no disclosures or conflicts of interest.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JPD-202472.

REFERENCES

[1] Surmeier DJ, Obeso JA, Halliday GM (2017) Selective neuronal vulnerability in Parkinson disease. *Nat Rev Neurosci* **18**, 101-113.

[2] Qian E, Huang Y (2019) Subtyping of Parkinson's disease - Where are we up to? *Aging Dis* **10**, 1130-1139.

[3] Centre for Evidence-Based Medicine (2018) Critical Appraisal of Prognostic Studies. Oxford University, Oxford. https://www.cebm.net/wp-content/uploads/2018/11/Prognosis.pdf

[4] Fereshtehnejad SM, Zeighami Y, Dagher A, Postuma RB (2017) Clinical criteria for subtyping Parkinson's disease: Biomarkers and longitudinal progression. *Brain* **140**, 1959-1976.

[5] Lawton M, Ben-Shlomo Y, May MT, Baig F, Barber TR, Klein JC, Swallow DMA, Malek N, Grosset KA, Bajaj N, Barker RA, Williams N, Burn DJ, Foltynie T, Morris HR, Wood NW, Grosset DG, Hu MTM (2018) Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J Neurol Neurosurg Psychiatry* **89**, 1279-1287.

[6] Landau S, Harris V, Burn DJ, Hindle JV, Hurt CS, Samuel M, Wilson KC, Brown RG (2016) Anxiety and anxious-depression in Parkinson's disease over a 4-year period: A latent transition analysis. *Psychol Med* **46**, 657-667.

[7] Zhang X, Chou J, Liang J, Xiao C, Zhao Y, Sarva H, Henchcliffe C, Wang F (2019) Data-driven subtyping of Parkinson's disease using longitudinal clinical records: A cohort study. *Sci Rep* **9**, 797.

[8] Erro R, Picillo M, Amboni M, Moccia M, Vitale C, Longo K, Pellecchia MT, Santangelo G, Martinez-Martin P, Chaudhuri KR, Barone P (2015) Nonmotor predictors for levodopa requirement in de novo patients with Parkinson's disease. *Mov Disord* **30**, 373-378.

[9] Lawton M, Baig F, Toulson G, Morovat A, Evetts SG, Ben-Shlomo Y, Hu MT (2020) Blood biomarkers with Parkinson's disease clusters and prognosis: The oxford discovery cohort. *Mov Disord* **35**, 279-287.

[10] Betrouni N, Delval A, Chaton L, Defebvre L, Duits A, Moonen A, Leentjens AFG, Dujardin K (2019) Electroencephalography-based machine learning for cognitive profiling in Parkinson's disease: Preliminary results. *Mov Disord* **34**, 210-217.

[11] Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intelle* **1**, 206-215.

[12] Schiess MC, Zheng H, Soukup VM, Bonnen JG, Nauta HJ (2000) Parkinson's disease subtypes: Clinical classification and ventricular cerebrospinal fluid analysis. *Parkinsonism Relat Disord* **6**, 69-76.

[13] Spiegel J, Hellwig D, Farmakis G, Jost WH, Samnick S, Fassbender K, Kirsch CM, Dillmann U (2007) Myocardial sympathetic degeneration correlates with clinical phenotype of Parkinson's disease. *Mov Disord* **22**, 1004-1008.

[14] Simuni T, Caspell-Garcia C, Coffey C, Lasch S, Tanner C, Marek K, Investigators P (2016) How stable are Parkinson's disease subtypes in de novo patients: Analysis of the PPMI cohort? *Parkinsonism Relat Disord* **28**, 62-67.

[15] Von Coelln FR, Barr E, Gruber-Baldini A, Reich S, Armstrong MJ, Shulman L (2015) Motor subtypes of Parkinson disease are unstable over time. *Neurology* **84**, S48.002.

[16] Rajput AH, Voll A, Rajput ML, Robinson CA, Rajput A (2009) Course in Parkinson disease subtypes: A 39-year clinicopathologic study. *Neurology* **73**, 206-212.

[17] Jankovic J, Kapadia AS (2001) Functional decline in Parkinson disease. *Arch Neurol* **58**, 1611-1615.

[18] Lee JW, Song YS, Kim H, Ku BD, Lee WW (2019) Alteration of tremor dominant and postural instability gait difficulty subtypes during the progression of Parkinson's disease: Analysis of the PPMI Cohort. *Front Neurol* **10**, 471.

[19] Selikhova M, Williams DR, Kempster PA, Holton JL, Revesz T, Lees AJ (2009) A clinico-pathological study of subtypes in Parkinson's disease. *Brain* **132**, 2947-2957.

[20] De Pablo-Fernandez E, Lees AJ, Holton JL, Warner TT (2019) Prognosis and neuropathologic correlation of clinical subtypes of Parkinson disease. *JAMA Neurol* **76**, 470-479.

[21] Kang JH, Mollenhauer B, Coffey CS, Toledo JB, Weintraub D, Galasko DR, Irwin DJ, Van Deerlin V, Chen-Plotkin AS, Caspell-Garcia C, Waligórska T, Taylor P, Shah N, Pan S, Zero P, Frasier M, Marek K, Kieburtz K, Jennings D, Tanner CM, Simuni T, Singleton A, Toga AW, Chowdhury S, Trojanowski JQ, Shaw LM (2016) CSF biomarkers associated with disease heterogeneity in early Parkinson's disease: The Parkinson's Progression Markers Initiative study. *Acta Neuropathol* **131**, 935-949.

[22] Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, Crespo I, Perumal TM, Jurkowski W, Antony PM, Diederich N, Buttini M, Kodama A, Satagopam VP, Eifes S, Del Sol A, Schneider R, Kitano H, Balling R (2014) Integrating pathways of Parkinson's disease in a molecular interaction map. *Mol Neurobiol* **49**, 88-102.

[23] Santiago JA, Bottero V, Potashkin JA (2017) Biological and clinical implications of comorbidities in Parkinson's disease. *Front Aging Neurosci* **9**, 394.

[24] Trezzi JP, Hiller K, Mollenhauer B (2018) The importance of an independent validation cohort for metabolomics biomarker studies. *Mov Disord* **33**, 856.

[25] Hepp DH, Vergoossen DL, Huisman E, Lemstra AW, Berendse HW, Rozemuller AJ, Foncke EM, van de Berg WD (2016) Distribution and load of amyloid-β pathology

in Parkinson disease and dementia with Lewy bodies. *J Neuropathol Exp Neurol* **75**, 936-945.

[26] Li YI, Wong G, Humphrey J, Raj T (2019) Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nat Commun* **10**, 994.

[27] Keo A, Mahfouz A, Ingrassia AMT, Meneboo JP, Villenet C, Mutez E, Comptdaer T, Lelieveldt BPF, Figeac M, Chartier-Harlin MC, van de Berg WDJ, van Hilten JJ, Reinders MJT (2020) Transcriptomic signatures of brain regional vulnerability to Parkinson's disease. *Commun Biol* **3**, 101.

[28] Shao Y, Le W (2019) Recent advances and perspectives of metabolomics-based investigations in Parkinson's disease. *Mol Neurodegener* **14**, 3.

[29] Weiner WJ (2008) There is no Parkinson disease. *Arch Neurol* **65**, 705-708.

[30] Titova N, Chaudhuri KR (2017) Personalized medicine in Parkinson's disease: Time to be precise. *Mov Disord* **32**, 1147-1154.